Data-driven lexical framing of ESG in infrastructure projects: an NLP-based decision support approach

Smart and Sustainable Built Environment

Mahshid Tootoonchy
Victoria University, Melbourne, Australia, and
Roksana Jahan Tumpa
School of Engineering and Technology, Central Queensland University, Sydney, Australia

Received 13 February 2026
Revised 12 May 2026
Accepted 3 July 2026

Abstract

Purpose – This study investigates how environmental, social and governance (ESG) priorities and digital transformation are linguistically framed in project-level discourse within sustainable infrastructure delivery. It examines how language shapes managerial attention, governance priorities and the interpretation of sustainability evidence, beyond formal indicators and dashboards. By applying natural language processing (NLP) to large-scale practitioner communications, the study aims to demonstrate how discourse-aware analytics can enhance transparency, reflexivity and balance in ESG-related decision-making and support more socially sensitive and context-aware project governance.

Design/methodology/approach – The study adopts a data-driven design using a large corpus of infrastructure project communications. A hybrid NLP pipeline is applied, combining topic modelling, semantic clustering and sentiment analysis to identify dominant ESG themes and evaluative patterns. To ensure interpretability and contextual validity, automated outputs are iteratively validated through human-in-the-loop expert review. The approach focuses on project-level discourse, enabling systematic analysis of how ESG issues, as well as digital technologies, are framed in practice to inform decision support and governance design.

Findings – The analysis reveals a pronounced asymmetry in ESG discourse within infrastructure projects. Environmental and governance themes are frequently framed in positive, data-driven terms and closely linked to digital tools such as dashboards, automation and monitoring systems. In contrast, social issues are less visible, more cautiously articulated and rarely digitally mediated, reflecting relational and normative framing. Sentiment analysis shows optimism associated with digitalized ESG narratives, suggesting a digitization bias that amplifies measurable dimensions while marginalizing social concerns. These findings indicate that language and digital mediation jointly shape managerial attention and may reproduce imbalances in ESG-oriented decision support.

Research limitations/implications – The study is limited by its sectoral focus on infrastructure projects, which may constrain generalizability to other industries. The use of lexicon-based sentiment analysis and unsupervised topic modelling may oversimplify nuanced or context-dependent ESG language, while potentially introducing algorithmic bias. In addition, the cross-sectional design captures a static snapshot of discourse rather than temporal evolution. Despite human-in-the-loop validation, some interpretive subjectivity remains in framing classification. Future research could extend this work through longitudinal designs, domain-specific NLP models and cross-sector validation to strengthen the robustness and generalizability of findings.

Practical implications – The study offers implications for infrastructure project governance and decision-support system (DSS) design. It demonstrates how ESG-related communication can be translated into actionable framing indicators, enabling more transparent and reflexive decision-making. Practically, organizations can integrate framing analytics into dashboards to complement traditional ESG metrics by visualizing sentiment, topic salience and digital framing patterns. This supports early detection of imbalances, such as the under-representation of social sustainability or over-reliance on digital optimism narratives. However, these tools should be used as decision-support aids rather than automated decision-making mechanisms, given the risk of misinterpretation of contextual and linguistic signals.

Originality/value – This study advances sustainable infrastructure research by introducing lexical framing as a form of "decision-metadata" that reveals how ESG priorities are interpreted and legitimized in project governance. Unlike prior ESG analytics that focus on indicators or predictive accuracy, it integrates NLP with

Smart and Sustainable Built Environment
Emerald Publishing Limited
e-ISSN: 2046-6102
p-ISSN: 2046-6099
DOI 10.1108/SASBE-02-2026-0122

interpretive, human-in-the-loop analysis to diagnose rhetorical asymmetries across ESG pillars. The study demonstrates how digital transformation shapes not only ESG measurement but also sustainability narratives, offering a novel, discourse-aware approach to decision support that enhances transparency, reflexivity and social balance in infrastructure project governance.

Keywords Sustainable infrastructure, ESG, Digital transformation, Lexical framing, Natural language processing

Paper type Research article

1. Introduction

Language plays a constitutive role in shaping how sustainability issues are understood, prioritized, and acted upon in project environments. It is not neutral; rather, it structures attention, assigns responsibility, and legitimizes trade-offs in decision-making processes (Weick, 1995; Cornelissen *et al.*, 2015). In project-based environments, where multiple actors with differing objectives interact, linguistic framing becomes central to defining problems and guiding collective action (Müller *et al.*, 2016; Too and Weaver, 2014). From a governance perspective, sustainability is therefore not only implemented through formal systems but continuously constructed through discourse that shapes what is considered decision-relevant (Flyvbjerg, 2014; Zwikael and Huemann, 2023).

This is particularly evident in sustainable infrastructure delivery, which has become a critical domain for addressing climate change, urbanization, and social inequality (Tumpa *et al.*, 2026). Infrastructure projects are increasingly expected to deliver not only economic value but also environmental protection, social inclusion, and ethical governance outcomes (Maurya and Biswas, 2021; Müller *et al.*, 2014). As a result, Environment (E), Social (S) and Governance (G) frameworks, initially indicator systems, have moved beyond corporate reporting tools to become embedded in project governance structures, supporting accountability and legitimacy across stakeholders (Eccles *et al.*, 2020; Wang *et al.*, 2025a, b). However, this expansion also increases the complexity of decision-making, as sustainability considerations must be justified across financial, technical, and societal dimensions simultaneously.

Infrastructure projects provide a distinctive empirical context compared to general ESG disclosure studies for several reasons. First, unlike firm-level ESG reporting, which is typically retrospective and strategically curated, infrastructure projects involve real-time, operational decision-making where sustainability priorities are continuously negotiated among multiple stakeholders (Müller *et al.*, 2016; Too and Weaver, 2014). Second, these projects are characterized by high complexity, long lifecycles, and multi-actor governance structures, requiring ongoing coordination between engineers, contractors, regulators, and communities (Flyvbjerg, 2014; Zwikael and Huemann, 2023). Third, ESG discourse in this context is embedded in internal communications (e.g. meetings, reports, digital platforms) rather than formal disclosures, making it less standardized and more reflective of actual decision processes. As a result, infrastructure projects offer a linguistically rich and underexplored setting in which ESG meaning is actively constructed, rather than merely reported. This enables the study to capture the micro-level framing dynamics that are typically obscured in aggregated corporate ESG datasets.

To manage this complexity, organizations increasingly rely on digital technologies and decision-support systems (DSS), including dashboards, artificial intelligence, building information modelling (BIM), and automated monitoring systems. These tools enhance transparency and enable real-time tracking of project performance (Abdel-Tawab *et al.*, 2023; Larbi *et al.*, 2026). However, existing research highlights a critical limitation: while such systems improve visibility, they also risk narrowing attention to what is measurable, potentially marginalizing interpretive and socially complex dimensions of sustainability (Lee *et al.*, 2025; Maibaum *et al.*, 2024). Consequently, governance failures often arise not from lack of data but from how information is framed, interpreted, and communicated in decision-making processes (Too and Weaver, 2014).

Within this context, framing theory offers a useful lens for understanding how ESG meanings are constructed in practice. Framing shapes how actors interpret sustainability information by influencing what is emphasized, what is omitted, and how trade-offs are justified (Entman, 1993; Benford and Snow, 2000). In DSS environments, framing also functions as a cognitive schema that shapes how information is attended to and how evidence is interpreted in decision-making (Arnott and Pervan, 2016). Despite its relevance, most ESG research focuses on external communication, such as corporate reports or public disclosures, where language is strategically curated to manage legitimacy (Boiral and Henri, 2017; Gałecka-Drozda *et al.*, 2021). This leaves a significant gap in understanding how ESG meaning is constructed within internal project-level communication.

In infrastructure projects, internal discourse among engineers, planners, contractors, and managers plays a critical role in translating ESG principles into action (Goh *et al.*, 2020). These interactions occur through meetings, reports, and digital platforms where sustainability objectives are continuously interpreted, negotiated, and adapted to project constraints (Qi *et al.*, 2023; Villa *et al.*, 2023). Yet, this intra-organizational layer of communication remains underexplored in ESG and DSS literature, despite being central to decision-making in complex project environments. As a result, existing studies provide limited insight into how sustainability priorities are actually constructed and operationalized during project delivery.

Recent advances in sustainability analytics and natural language processing (NLP) have enabled new ways of analyzing ESG discourse. However, much of this work remains focused on structured corporate disclosures and external reporting datasets. While these approaches improve scalability and classification accuracy, they often overlook the interpretive and framing processes embedded in practitioner communication (Maibaum *et al.*, 2024; Roufosse *et al.*, 2024). This limitation is particularly significant in infrastructure contexts, where language is not only descriptive but also performative, shaping legitimacy and decision outcomes.

Against this background, this study is situated within the sustainable built environment and infrastructure governance literature. Unlike firm-level ESG studies that focus on reporting and disclosure quality, infrastructure projects represent dynamic environments where sustainability is actively constructed through ongoing communication among multiple stakeholders (Weerakoon and Perera, 2025). These projects are characterized by long lifecycles, complex governance structures, and competing environmental, social, and economic objectives. This makes them a distinctive empirical context for examining how ESG meaning is constructed through language during project delivery.

Building on this context, the study integrates framing theory with DSS and NLP-based text analytics to examine how ESG priorities and digital transformation are linguistically constructed in infrastructure project communication. It argues that practitioner language is not merely descriptive but constitutive: it shapes what is measured, prioritized, and legitimized in decision-making processes. This aligns with DSS research, which emphasizes that analytical artefacts influence decision framing as much as they inform decisions (Arnott and Pervan, 2016; Marques and Ferreira, 2020).

Accordingly, this study addresses the following research question:

RQ. How do ESG priorities and digital transformation get linguistically framed in project discourse, and how can NLP-based analysis inform sustainable project governance and decision support?

To address this question, the study is guided by four sub-questions:

RQ1. What topics and themes emerge most frequently across ESG discourse?

RQ2. How are ESG priorities emphasized or downplayed through language?

RQ3. Are there discernible differences in lexical framing between E, S, and G components?

RQ4. How do digital technologies get positioned in relation to ESG themes?

Using a large corpus of professional communications, we apply a hybrid NLP pipeline (topic modelling, semantic clustering, sentiment analysis) with human-in-the-loop validation, following best practices in explainable DSS design (Lee *et al.*, 2025). This approach moves beyond keyword detection to reveal what is highlighted, omitted, and how digital references influence evaluative tone.

Our study conceptualizes lexical framing as "decision metadata" supporting reflexive governance. Practically, DSS can integrate framing diagnostics to ensure transparency, contextual sensitivity, and social balance. The study contributes to sustainable built environment research by shifting the focus from ESG *measurement* to ESG *meaning*. It demonstrates that sustainability in infrastructure projects is not only quantified through digital systems but also constructed through linguistic framing processes that influence decision-making, governance, and legitimacy. In doing so, it advances an interpretive DSS perspective in which understanding how sustainability is communicated is as important as measuring sustainability outcomes. Unlike explainable AI, which clarifies model predictions, or semantic features used for classification, decision metadata refers to higher-order linguistic attributes (e.g. framing, tone, and digital coupling) that characterize how sustainability issues are constructed and prioritized within discourse, thereby informing how decision contexts are structured rather than predicting decisions themselves (Know *et al.*, 2009).

The remainder of the paper: Section 2 reviews ESG integration, digital transformation, and framing theory. Section 3 details data, NLP methods, and validation. Section 4 presents thematic, sentimental, and digital framing analyses. Section 5 reports findings, followed by Section 6 on implications, governance, and future research agenda.

2. Literature review

2.1 ESG in infrastructure projects: context and problem definition

Infrastructure projects lie at the intersection of economic growth, environmental responsibility, and social legitimacy (Akomea-Frimpong *et al.*, 2025; Tumpa *et al.*, 2025). They are capital-intensive, politically visible, and technologically complex, making financing, design, and operational decisions inherently tied to ESG considerations (Ansar *et al.*, 2016; Wang *et al.*, 2025a, b). Global sustainability agendas, such as the UN SDGs and the Paris Agreement, have reframed infrastructure as a central vehicle for low-carbon, inclusive development.

Governance operates under persistent uncertainty. Projects must satisfy diverse stakeholders (governments, investors, regulators, communities) while coping with regulatory volatility and geopolitical risk (Chan *et al.*, 2022; Poorisat *et al.*, 2026). Decision-making is rarely supported by integrated analytical systems that balance financial performance with ESG impacts, motivating interest in data-driven DSS to synthesize heterogeneous sustainability information and enhance transparency (Burstein and Holsapple, 2008; Arnott and Pervan, 2016).

ESG disclosure began as a governance innovation rather than a computational task. The UN Global Compact (2006) and Principles for Responsible Investment (Gasperini, 2019) framed ESG as a means to integrate ethical and sustainability concerns into corporate accountability. Regulatory mandates in Europe, the UK, and Asia have institutionalized ESG reporting (Gong *et al.*, 2024; Gao *et al.*, 2025). Yet, unlike financial reporting, ESG remains plural, narrative-driven, and inconsistent, complicating comparability and integration into analytic models (Boiral and Henri, 2017; Van Tam and Toan, 2026).

Infrastructure projects magnify these challenges. The sector contributes nearly 39% of global CO_2 emissions and consumes about one-third of natural resources. Despite this footprint, firms often lack data infrastructures and analytical capacity for systematic ESG management (Debrah *et al.*, 2022; Li *et al.*, 2021). Reporting is frequently outsourced, expensive, and descriptive rather than diagnostic (Ure *et al.*, 2024), so what is measured often reflects strategic image management more than operational practice, biasing stakeholder judgement and algorithmic assessment.

Text data offers an under-exploited, decision-relevant resource. Project correspondence, meeting transcripts, and digital platform exchanges contain first-hand reflections of practitioners' priorities, doubts, and innovations. Computational analysis of these texts exposes the "raw" informational substrate that conventional ESG ratings overlook, revealing how sustainability is interpreted in practice. Our first research sub-question is: RQ1: What topics and themes emerge most frequently across ESG discourse in infrastructure projects? Understanding these themes provides the foundation for designing DSS pipelines that translate practitioner language into actionable sustainability intelligence.

2.2 Prior research on ESG reporting and assessment

Existing research on ESG assessment is extensive but fragmented across accounting, sustainability management, and information systems domains. Established frameworks such as the Global Reporting Initiative (GRI), Integrated Reporting (IR), and Sustainability Accounting Standards Board Standards have improved non-financial disclosure, yet empirical studies consistently identify issues of selective emphasis, symbolic compliance and limited comparability (Baldini *et al.*, 2018; Boiral and Henri, 2017). From a DSS standpoint, these limitations extend beyond reporting quality to analytical usability as inconsistent indicators and narrative bias to reduce the machine-readability of ESG data and hinder automated decision-making.

Recent DSS-oriented studies have begun to address these challenges through text analytics and computational methods. Maibau *et al.* (2024) demonstrate that ESG classification outcomes are highly sensitive to methodological choices, including preprocessing and tool selection, indicating that ESG analytics are inherently shaped by analytical framing rather than being neutral representations of sustainability. Similarly, Lee *et al.* (2025) propose an NLP-based ESG-KIBERT framework that improves classification accuracy and interpretability; however, their focus remains on structured corporate disclosures rather than the informal, practitioner-generated communication that characterizes project environments. In parallel, AI-enabled decision-support systems are increasingly being embedded across procurement and governance processes to enhance visibility and efficiency (Lakhal *et al.*, 2025).

Accounting and Project Management research has examined organizational drivers of ESG disclosure. For example, Radzi *et al.* (2026) found that stronger governance mechanisms correlate with higher ESG transparency, whereas Gałecka-Drozda *et al.* (2021) noted that self-reporting often overstates achievements, a persistent "greenwashing" tendency echoed in later studies. Despite these advances, limited attention has been given to the micro-level language through which ESG meanings are constructed, particularly in project-based settings where decisions are negotiated rather than formally reported.

Within the infrastructure and built environment literature, existing studies remain scarce. Previous research explored ESG adoption barriers in construction supply chains (Debrah *et al.*, 2022), the role of Building Information Modelling (BIM) in environmental tracking (Abdel-Tawab *et al.*, 2023). While these studies highlight digital tools in enabling sustainability outcomes, they rarely analyze the communication processes or interpretive framing through which ESG priorities are articulated and justified in practice. Moreover, infrastructure projects differ from firm-level ESG contexts due to their long lifecycles, multi-stakeholder governance structures, and complex trade-offs between environmental, social, and economic objectives. To clarify the evolution of this literature and the analytical space our study occupies, Table 1 summarizes representative works across sustainability, ESG, and project domains.

Across these studies, three key gaps emerge. First, most studies rely on curated or formalized reporting data, which reflects strategic representation rather than operational decision-making discourse. Second, the methodological emphasis remains on classification

Table 1. Prior studies on ESG assessment and reporting

Study	Domain/Data	Methodology	Key findings	Limitations/Relevance to this study
Boiral (2009)	Corporate reports (mining sector)	Manual content analysis	Revealed selective emphasis and greenwashing in disclosures	Small sample: qualitative bias, shows need for scalable text analytics
Baldini *et al.* (2018)	Cross-industry firms	Panel regression on GRI adoption	Institutional pressures drive ESG disclosure	Quantitative but no textual depth
Maibaum *et al.* (2024)	Corporate ESG reports	Comparative text-mining tool evaluation	Tool choice alters classification accuracy and interpretability	Highlights the importance of methodological transparency for DSS design
Lee *et al.* (2025)	Cross-sector ESG data	NLP model (ESG-KIBERT)	Improved classification using domain lexicons	Focuses on scoring; no discourse context
Abdel-Tawab *et al.* (2023)	Construction projects	Case study on BIM implementation	BIM improves environmental sustainability tracking	Technological focus; lacks linguistic analysis
Gałecka-Drozda *et al.* (2021)	Corporate ESG reports	Content and sentiment analysis	Detected greenwashing patterns	External disclosure only
Qi *et al.* (2023)	Infrastructure projects	Mixed methods case study	Identified implementation barriers to ESG integration	Lacks computational text analysis
Roufosse *et al.* (2024)	Corporate texts	Knowledge-aware transformer model	Enhanced ESG classification accuracy	Predictive orientation: ignores framing effects

performance measurement rather than interpretive insight into how ESG meaning is constructed, negotiated, and framed. Third, infrastructure projects are rarely examined as linguistically rich decision environments, despite their central role in sustainability transitions. Addressing these gaps, our research shifts the focus from *"how to score ESG"* to *'how ESG is linguistically constructed in real-world project communication'*. By doing so, the study responds to calls for explainable, human-in-the-loop analytical approaches that integrate computational methods with interpretive understanding (Shollo and Galliers, 2016; Arnott and Pervan, 2016). Therefore, analyzing practitioner discourse through NLP reveals thematic priorities and blind spots that traditional ESG assessment methods overlook.

2.3 Digital transformation and ESG: data-driven governance or discursive bias

Digital transformation is now inseparable from ESG practice, especially in infrastructure projects with complex supply chains, fragmented accountability, and information asymmetry (Hosseini *et al.*, 2026; Najafi *et al.*, 2025). Technologies such as BIM, blockchain, IoT sensors, and AI analytics are widely positioned as enablers of transparency, efficiency, and environmental monitoring (Abdel-Tawab *et al.*, 2023; Peng *et al.*, 2023). From a DSS perspective, these technologies integrate operational data with managerial decision logic, shifting sustainability governance from periodic reporting to continuous data-driven monitoring (Marques and Ferreira, 2020; Power, 2022).

However, digitalization does not automatically translate into improved sustainability outcomes (Singh and Kumar, 2024). Rather, it reshapes what is made visible and measurable within ESG systems. Dashboards and digital ledgers tend to privilege quantifiable indicators, reinforcing a form of visibility bias that elevates environmental and governance metrics while marginalizing socially complex issues such as labour conditions and community impacts (Lee *et al.*, 2025; Shollo and Galliers, 2016). In this sense, digital infrastructures not only support ESG assessment but also actively shape what is constructed as ESG-relevant.

Empirical studies reinforce this concern. Maibaum *et al.* (2024) demonstrate that ESG classification outcomes vary depending on analytical configuration and preprocessing choices, indicating that computational ESG analysis is inherently shaped by framing effects. Similarly, explainable AI research emphasizes human oversight in algorithmic ESG scoring to avoid institutional blind spots (Roufosse *et al.*, 2024). Together, these studies suggest that digital systems embed interpretive assumptions that influence sustainability representation.

Within infrastructure projects, digital tools function not only as analytical systems but also as communicative and legitimizing (Ramohlokoane *et al.*, 2025). For example, BIM supports carbon tracking and waste monitoring, enabling evidence-based environmental reporting (Abdel-Tawab *et al.*, 2023; Mandičák *et al.*, 2024). Similarly, practitioners often invoke digitalization adoption as evidence of innovation or sustainability commitment, even when substantive outcomes are uncertain. This reflects a "technological optimism" frame, where the presence of digital systems is interpreted as progress in itself (Su *et al.*, 2023).

As shown in Table 2, Environmental and Governance clusters exhibit higher sentiment scores and stronger digital mediation compared to social clusters, providing quantitative support for the observed asymmetry.

Recent reviews highlight a paradox: while digitalization enables real-time ESG monitoring, it can also narrow organizational attention to what is technically measurable, excluding qualitative and relational dimensions of sustainability DSS research therefore emphasizes not only automation and accuracy but also interpretability; arguing that systems must explain how insights are generated to support responsible decision-making (Arnott and Pervan, 2016; Marques and Ferreira, 2020). To situate these debates, Table 2 compares key studies on digital transformation in ESG and DSS contexts.

Based on the review, three patterns emerge. First, digital tools are primarily framed as enablers of visibility and accountability, yet their role in shaping ESG discourse remains underexplored.

Table 2. Digital transformation and ESG (2020–2025)

Study	Digital focus/ Domain	Methodology	Main findings	Limitations/Relevance to this study
Abdel-Tawab *et al.* (2023)	BIM in construction	Case study	BIM improves environmental tracking and waste management	Demonstrates digital enablement but lacks linguistic/framing analysis
Kouhizadeh and Sarkis (2018)	Blockchain supply chains	Empirical and conceptual	Blockchain enhances transparency, mitigates social risk	Technological focus; no study of discourse or meaning construction
Power (2022)	Intelligent auditing	Conceptual review	Decision automation benefits compliance monitoring	Calls for integrating human judgement support the need for hybrid DSS
Maibaum *et al.* (2024)	Text-analysis tools	Comparative evaluation	Tool choice alters sustainability classification results	Shows methodological reflexivity; motivates explainable NLP design
Roufosse *et al.* (2024)	Transformer models for ESG	Deep learning study	Domain-tuned model improves ESG classification	Predictive focus: ignores rhetorical or contextual dimensions
Su *et al.* (2023)	Stakeholder digital engagement	Survey/SEM	Digital tools improve stakeholder satisfaction	Examines perception outcomes, not communicative framing

Second, social sustainability remains digitally under-represented. While environmental and governance aspects benefit from dashboards and automated compliance, labour ethics and community relations rely on narratives resistant to quantification (Villa *et al.*, 2023).

Third, little attention is given to how digital terminology in practitioner communication reflects underlying framing processes that influence legitimacy and interpretation. This informs RQ4: How are digital technologies positioned relative to ESG themes? Addressing this question allows the study to distinguish when digital artefacts function as analytical tools versus rhetorical signals of sustainability legitimacy, thereby improving understanding of framing effects in DSS-supported ESG assessment.

2.4 Communication, governance, and framing: from external signalling to internal sensemaking

Farming provides an interpretive and linguistic mechanism through which organizations construct meaning around ESG. It shapes what aspects of sustainability are emphasized, how problems, responsibilities, and achievements are interpreted. Entman (1993) conceptualizes framing as selecting salient aspects of reality to promote specific problem definitions, causal interpretations, moral evaluations, and treatment recommendations. Benford and Snow (2000) extend this to collective action, describing diagnostic, prognostic, and motivational framing as stages aligning sensemaking and mobilization. In a DSS context, framing functions as a decision schema that shapes attention, interpretation of uncertainty, and what is treated as valid evidence.

Most ESG-framing studies focus on external disclosures (reports, briefings, media). These studies show that organizations often emphasize positive outcomes while downplaying trade-offs, contributing to selective representation and greenwashing concerns (Boiral and Henri, 2017; Gałecka-Drozda *et al.*, 2021). However, this work offers limited insight into how ESG meaning is constructed during day-to-day decision-making, which is critical for DSS that rely on authentic, context-rich data (Shollo and Galliers, 2016).

In infrastructure projects, internal or practitioner-levels among engineers, planners, and contractors play a key role in interpreting and operationalizing ESG goals. They engage in continuous dialogue via meetings, reports, and digital platforms, negotiating objectives, interpreting sustainability directives, and resolving ethical dilemmas (Qi *et al.*, 2023; Villa *et al.*, 2023). Yet, this intra-organizational framing remains underexplored despite its importance for understanding ESG enactment in practice.

Sustainability accounting research has largely focused on legitimacy and disclosure motives (Del Gesso and Lodhi, 2025), but it explains why organizations report ESG rather than how the meaning is constructed. Legitimacy itself is discursive, negotiated through languages (Cornelissen *et al.*, 2015), yet micro-linguistic mechanisms such as tone and emphasis remain underexamined in project contexts.

ESG framing varies across pillars: environmental discourse adopts technical and quantitative vocabulary (emission, efficiency, tracking, dashboard), social discourse uses ethical and relational language (community, trust, equity), and governance discourse is procedural (audit, compliance, policy, system). These lexical asymmetries reflect distinct epistemic logics. Automated analytics risk over-valuing measurable dimensions (E and G) while marginalizing nuanced social aspects, reproducing structural imbalances in evaluation (Lee *et al.*, 2025; Maibaum *et al.*, 2024). Empirical evidence shows that framing influences ESG interpretation and perceived credibility (Kallioupi and Triantafyllou, 2025), yet prior work remains focused on external perception rather than internal decision processes. The missing layer is how practitioners construct and reinterpret ESG meaning during project delivery.

Overall, internal framing reveals a tension between technological optimism, where data-driven language signals progress, and ethical reflection, where social issues are framed as risks or constraints. Capturing these contrasts allows DSS to differentiate compliance from genuine engagement. RQ2 asks how ESG priorities are emphasized or downplayed; RQ3 examines differences in lexical framing across ESG pillars, linking micro-linguistic variation to macro-governance outcomes.

Methodologically, framing analysis bridges communication studies and computational text mining. Emerging hybrid designs integrate linguistic context into algorithmic classification (Roufosse *et al.*, 2024), reflecting that effective DSS depends on both data quantity and discourse quality. Quantifying frames (sentiment, modality, digital-vocabulary co-occurrence) allows NLP to operationalize qualitative insights at scale, aligning outputs with governance concerns.

2.5 Methods of studying ESG language

Research on ESG communication has traditionally relied on interpretive approaches such as thematic coding, discourse analysis, and qualitative content analysis. These methods offer depth and contextual insights into moral reasoning and rhetorical nuance in sustainability narratives (Boiral, 2009; Unerman and Chapman, 2014). However, they are limited in scalability and replicability in infrastructure projects where large volumes of meeting notes, audit reports, and stakeholder exchanges are generated. Manual approaches struggle to process the volume and heterogeneity of text. This creates a methodological gap for DSS that require both scale and interpretability.

To address these limitations, early computational attempts introduced dictionary-based sentiment, keyword extraction, and topic modelling. Studies such as Kiriu and Nozaki (2020) and Perazzoli *et al.* (2022) applied NLP techniques to ESG-related texts improving scalability and objectivity. However, these approaches often reduce sustainability discourse to frequency-based representations, limiting interpretive depth and overlooking contextual meaning. They also assume alignment between reported language and organizational behaviour, despite growing evidence of strategic disclosure bias and greenwashing detection (Gałecka-Drozda *et al.*, 2021; Lee *et al.*, 2025), particularly in project-based and infrastructure settings.

Recent research has moved toward hybrid and explainable NLP approaches that combine computational efficiency with interpretive validity. Maibaum *et al.* (2024) demonstrate that algorithmic transparency (including preprocessing, lexicon design, and model choice) is critical for decision credibility. Roufosse *et al.* (2024) introduce ESG-KIBERT, a knowledge-aware transformer that integrates domain ontologies to improve semantic coherence. Yet these approaches still prioritize predictive accuracy and rarely capture rhetorical structures such as emphasis, hedging, or evaluative polarity that shape how ESG is framed in professional discourse on projects.

A key limitation across existing approaches is the weak capture of framing-related linguistic features such as emphasis, hedging, and evaluative tone, which are central to how ESG meaning is constructed in organizational settings. Shollo and Galliers (2016) argue that effective decision support depends not only on analytics but on how outputs are interpreted within organizational narratives. To address this challenge, recent methodological developments advocate multi-stage NLP pipelines combining topic modelling (Blei *et al.*, 2003), embedding-based clustering (Jain, 2010), sentiment analysis (Hutto and Gilbert, 2014), and keyword extraction (Rose *et al.*, 2010; Maibaum *et al.*, 2024). When integrated within an auditable workflow, these methods support DSS objectives of transparency, traceability, and iterative learning (Arnott and Pervan, 2016).

Despite these advances, most NLP-based ESG studies remain focused on corporate disclosures such as annual reports or sustainability websites that overlook internal project communication, where ESG meaning is actively constructed and negotiated. In infrastructure projects, where multidisciplinary teams negotiate ESG trade-offs under time, contractual, and regulatory constraints, capturing practitioner language is essential for DSS that reflect real-world sensemaking in construction and infrastructure delivery.

Interpretability presents an additional concern. While deep-learning models achieve high accuracy, they often operate as "black boxes." In ESG governance, such opacity is problematic, as stakeholders must understand why a statement is classified as positive, sceptical, or risk-oriented. Explainable AI (XAI) approaches, therefore, emphasize techniques such as attention visualization and feature attribution (Marques and Ferreira, 2020). Embedding these features within NLP pipelines supports accountability and trust, key requirements for sustainability governance in the built environment.

Overall, this study positions NLP as a decision-support mechanism rather than a purely predictive tool. It contributes by (1) extending text mining from corporate disclosures to practitioner-level communications within infrastructure projects; (2) operationalizing framing constructs such as tone, emphasis, and digital cues as measurable variables; and (3) embedding explainability and human validation throughout the analytic pipeline.

Three key challenges remain in literature. First, domain adaptation remains limited as ESG language varies across sectors, and generic models often misclassify infrastructure-specific expressions (Lee *et al.*, 2025). Second, cross-lingual and cross-cultural variation in ESG discourse is underexplored, despite the multinational nature of infrastructure projects (Marques and Ferreira, 2020). Third, qualitative validation is rarely integrated into computational workflows, despite DSS literature emphasizing iterative human interpretation as essential for system credibility (Arnott and Pervan, 2016). This research addresses these gaps by embedding manual review and coder triangulation within the NLP process.

2.6 Synthesis of gaps and conceptual positioning

The preceding review reveals four key gaps in existing ESG and DSS literature that motivate this study. These gaps arise from limitations in how ESG is currently conceptualized, measured, and analyzed, and they directly inform the study's sub-research questions (Table 3). Conceptually, the study sits at the intersection of ESG governance, digital transformation, and decision-support analytics. It argues that practitioner language is not merely descriptive but constitutive: the way ESG is discussed shapes what is measured, prioritized, and legitimized. This aligns with DSS research, which emphasizes that analytical artefacts influence decision framing as much as they inform decisions (Marques and Ferreira, 2020; Arnott and Pervan, 2016).

Building on this perspective, this study extends ESG analytics in project management from predictive classification towards interpretive diagnostics. While existing approaches primarily quantify sustainability outcomes, this research focuses on explaining how such outcomes are constructed through language and framing, contributing to an explainable paradigm emphasizing transparency, interpretability, and contextual grounding (Maibaum *et al.*, 2024; Lee *et al.*, 2025).

Theoretically, the study advances the concept of "decision metadata", where linguistic features such as tone, framing, and digital vocabulary are treated as observable indicators of discursive and institutional patterns embedded in ESG communication. Unlike explainable AI or semantic classification approaches, which focus on prediction and model transparency, decision metadata captures how sustainability issues are constructed, prioritized, and made visible within discourse. This transforms language into a diagnostic layer of decision support, enabling the identification of framing asymmetries, digital visibility biases, and interpretive gaps that are not captured by conventional ESG analytics.

To operationalize framing computationally, this study maps established framing theory constructs (Entman, 1993; Benford and Snow, 2000) to observable linguistic features within the NLP pipeline. Specifically, topic modelling captures problem definition by identifying dominant ESG themes, while sentiment analysis reflects evaluative orientation (e.g. positive, neutral, or critical framing). Clustering and keyword co-occurrence analysis capture salience and emphasis by identifying recurrent lexical patterns and their relative prominence.

Table 3. Summary of research gaps and corresponding sub-research questions

Identified research gap	Description	Corresponding RQ(s)	Implications for DSS
Gap 1: Topic visibility in practitioner discourse	Project-level ESG communication is fragmented and filtered through strategic external disclosures; practitioner narratives remain under-analyzed	RQ1: What topics and themes emerge most frequently across ESG discourse in infrastructure projects?	Highlights the need for DSS pipelines capable of mining operational texts to reveal latent sustainability priorities
Gap 2: Positioning and bias of digital transformation	Digital tools are often celebrated as enablers of sustainability, yet little is known about how they are linguistically framed or whether this reflects substantive outcomes	RQ4: How are digital technologies positioned in relation to ESG themes?	Supports the development of DSS models that detect optimism bias and distinguish rhetorical from functional digitalization
Gap 3: Internal framing and pillar asymmetry	Most studies focus on external ESG reports; few examine how practitioners internally frame environmental, social, and governance issues differently	RQ2 and 3: How are ESG priorities emphasized or downplayed through language?/Are there discernible differences in lexical framing between E, S, and G components?	Encourages DSS designs sensitive to linguistic and affective asymmetries across ESG pillars
Gap 4: Computational-interpretive integration	Existing NLP models privilege predictive accuracy over interpretive insight; qualitative meaning is seldom linked to algorithmic output	RQ1 – 3	Necessitates explainable, human-in-the-loop DSS that combine scalable computation with discourse interpretation

In addition, the co-occurrence of digital terminology (e.g. “dashboard”, “AI”, “automation”) with ESG topics is treated as an indicator of technological framing, reflecting how digital mediation shapes the presentation of sustainability issues. Finally, subjectivity scores provide a proxy for rhetorical intensity, distinguishing between descriptive and evaluative discourse. Together, these features extend framing analysis beyond sentiment polarity, enabling a multidimensional and computationally tractable representation of how ESG issues are constructed in project discourse. To enhance transparency and theoretical integration, Table 4 presents the mapping between framing theory constructs and their computational operationalization within the NLP pipeline.

3. Method

3.1 Data collection and preprocessing

The dataset comprises a large, multi-source corpus of professional communications related to infrastructure projects, including meeting transcripts, technical documentation, and digital platform interactions. This diversity ensures representation of operational, managerial, and governance-level language, providing a realistic linguistic substrate for decision-support interpretation.

To ensure data quality and analytic rigour, the texts underwent extensive cleaning and preprocessing following established NLP protocols. Steps included tokenization, stop word removal, lemmatization, and lowercasing (Brachem and Rothe, 2021). The spaCy library (Honnibal and Montani, 2017) was used for efficient linguistic annotation and lemmatization, enabling standardized text normalization conducive to subsequent analysis.

Table 4. Operationalization of framing constructs in the NLP pipeline

Framing dimension	Theoretical basis	NLP operationalization	Analytical technique	Interpretive role in this study
Problem definition	Entman (1993)	Identification of dominant ESG themes and issues	Topic modelling (LDA) + clustering	Captures how sustainability issues are *defined and categorized* in project discourse
Diagnostic framing	Benford and Snow (2000)	Detection of negative or neutrally framed issues (e.g., risk, compliance challenges)	Sentiment analysis (VADER compound, polarity scores)	Indicates how ESG issues are *problematized or contested*
Prognostic framing	Benford and Snow (2000)	Positive framing linked to solutions, innovation, or improvement	Sentiment analysis + co-occurrence with digital keywords	Reflects how ESG issues are *positioned as solvable through actions or technologies*
Motivational framing	Benford and Snow (2000)	Degree of evaluative intensity and rhetorical emphasis	Subjectivity scores (TextBlob) + sentiment strength	Captures how discourse *encourages, justifies, or legitimizes action*
Salience/ Emphasis	Entman (1993)	Frequency and recurrence of keywords and topics within clusters	Keyword frequency analysis + cluster density	Identifies which ESG issues are *prioritized or made more visible*
Lexical framing	Framing theory + ESG discourse literature	Distinct vocabulary patterns across E, S, G (e.g., technical vs relational vs procedural language)	Keyword extraction + co-occurrence analysis	Reveals *how ESG dimensions differ linguistically and conceptually*
Digital framing (Technological mediation)	DSS and digital transformation literature (Marques and Ferreira, 2020; Power, 2022; Maibaum *et al.*, 2024)	Co-occurrence of ESG topics with digital terms (e.g., AI, dashboard, automation)	Keyword dictionary + contextual filtering	Captures how digital technologies *shape the presentation and perceived legitimacy of ESG issues*
Framing intensity/Tone	Entman (1993)	Overall evaluative polarity and variation across ESG dimensions	Aggregated sentiment statistics + comparative tests (Kruskal–Wallis)	Enables *cross-dimensional comparison* of ESG discourse tone

To meet reproducibility and explainability standards, preprocessing was implemented through fully auditable Python scripts, each step automatically logged and timestamped. This ensured traceability of all transformations and supported transparency for subsequent model interpretation, which is a core expectation in responsible decision-support environments (Provost and Fawcett, 2013; Power, 2022).

3.2 Natural language processing techniques

To explore the lexical framing of ESG priorities within infrastructure discourse, we employed a multi-layered NLP pipeline combining unsupervised and lexicon-based methods. This hybrid design follows recent recommendations for "explainable analytics," integrating interpretability with scalability (Lee *et al.*, 2025; Maibaum *et al.*, 2024; Marques and Ferreira, 2020).

The analytical stages included:

(1) Topic Modelling: Latent Dirichlet Allocation (LDA) (Blei *et al.*, 2003) was applied to extract latent thematic structures from the corpus, revealing dominant ESG-related themes across environmental, social, and governance dimensions. Implementation was through Gensim (Řehůřek and Sojka, 2010), allowing scalable, interpretable topic extraction from large unstructured text corpora. While more recent embedding-based topic models (e.g. transformer-based clustering approaches) provide enhanced contextual representation, LDA was selected due to its high interpretability and transparent probabilistic structure, which is critical for decision-support applications where traceability of lexical contributions is required (Arnott and Pervan, 2016). This aligns with the study's focus on lexical framing as an interpretable form of decision-relevant metadata.

(2) Semantic Clustering: To capture relational proximity among discourse segments, document embeddings were generated using Term Frequency–Inverse Document Frequency (TF-IDF) vectors, followed by dimensionality reduction via Principal Component Analysis (PCA). The reduced vectors were clustered with the k-means algorithm to group semantically similar discourse segments (Jain, 2010). This process enabled pattern discovery across ESG topics and project contexts, consistent with DSS goals of structuring unstructured knowledge (Shang *et al.*, 2023).

(3) Key-Phrase Extraction: The Rapid Automatic Keyword Extraction (RAKE) algorithm (Rose *et al.*, 2010) identified salient ESG-related terminology based on syntactic and frequency heuristics. This step generated interpretable lexical anchors for managerial sensemaking, supporting the transparency dimension of DSS explainability (Shollo and Galliers, 2016).

(4) Sentiment Analysis: To capture affective aspects of ESG framing, sentiment polarity was measured using the VADER lexicon-based tool (Hutto and Gilbert, 2014). VADER's fine-grained treatment of negation and intensity enabled differentiation of positive, neutral, and negative tones, reflecting stakeholder emotions embedded in textual discourse. This emotional layer provides additional decision metadata (how practitioners emotionally orient to ESG priorities), expanding the interpretive range of DSS analytics. Although domain-specific and deep learning-based sentiment models may achieve higher predictive accuracy, they often operate as black-box systems or require extensive labelled data. In contrast, VADER provides transparent, lexicon-based scoring that supports interpretability. To address domain sensitivity in ESG discourse, sentiment outputs were cross validated through manual coding and expert review, ensuring alignment with project governance language.

(5) Digital Framing Detection: To assess how digital transformation is positioned linguistically within ESG narratives, we systematically detected co-occurrence of digital technology terms (e.g. AI, blockchain, automation, dashboard). This stage reveals how technological vocabulary intersects with ESG discourse and indicates where digitalization frames sustainability as efficiency, control, or innovation (van der Velden, 2018.). Digital framing was operationalized through keyword-based detection of technology-related terms, including AI, artificial intelligence, machine learning, automation, dashboard, platform, system, digital, data analytics, blockchain, sensor, monitoring, and integration.

(6) To address ambiguity (e.g. "system," "platform"), terms were only retained when co-occurring with ESG-relevant context within the same sentence or segment. Ambiguous cases were further reviewed during the human-in-the-loop validation stage to ensure contextual accuracy.

(7) A cluster was classified as "digitally mediated" when at least 20% of its documents contained one or more validated digital keywords, consistent with threshold-based approaches in text classification and DSS-oriented analytics.

3.3 Computational tools and implementation

All analyses were implemented in Python 3.9 within a Jupyter Notebook environment, facilitating iterative exploration and reproducibility. The following libraries were utilized:

(1) spaCy for text preprocessing (Honnibal and Montani, 2017)

(2) Gensim for topic modelling (Řehůřek and Sojka, 2010)

(3) Scikit-learn for clustering (Pedregosa *et al.*, 2011)

(4) NLTK for text normalization and stopword management (Bird *et al.*, 2009)

(5) RAKE-NLTK for key phrase extraction (Rose *et al.*, 2010)

(6) VADER Sentiment Analyzer for sentiment scoring (Hutto and Gilbert, 2014)

Each component was modularized into a traceable workflow, allowing parameter adjustment and visualization at each stage (see Table 5). This approach is consistent with methodological standards for human–system interaction and interpretive transparency (Arnott and Pervan, 2016; Power, 2022).

This pipeline summarizes the sequential transformation from raw textual data to interpretable ESG framing indicators, supporting transparency and reproducibility in the analytical workflow. Some clusters contain terms that may appear contextually peripheral (e.g. "today"). These arise from the use of real-world project communication data, where operational or incidental language co-occurs with ESG-related discourse. While preprocessing steps removed standard stop words, domain-specific noise may persist due to contextual proximity within documents.

Importantly, cluster interpretation is based on dominant semantic patterns rather than isolated keywords, and all clusters were reviewed during the human-in-the-loop validation stage to ensure relevance and interpretability.

3.4 Validation and reliability

To ensure robustness, multiple validation procedures were implemented:

(1) Topic coherence was assessed using the coherence score (Röder *et al.*, 2015) to determine the optimal number of topics and confirm interpretability. Topic modelling

Table 5. The pipeline summary

Step	Method	Purpose	Output
Preprocessing	Tokenization, lemmatization (spaCy)	Clean and normalize text	Standardized corpus
Topic modelling	LDA (Gensim)	Identify latent themes	Topic distributions
Clustering	TF-IDF + PCA + k-means	Group similar discourse segments	Semantic clusters
Key-Phrase extraction	RAKE	Extract salient ESG terms	Keyword sets
Sentiment analysis	VADER	Measure evaluative tone	Sentiment scores
Digital framing	Keyword detection	Identify digital mediation	Binary/cluster-level tagging
Validation	Coherence, silhouette, expert review	Ensure robustness	Validated outputs

quality was assessed using the coherence score (C_v), with the selected model achieving a value of approximately 0.40, indicating acceptable semantic interpretability. The moderate coherence reflects the use of real-world conversational data, where thematic overlap and contextual variability are expected.

(2) Clustering quality was evaluated using silhouette coefficients (Rousseeuw, 1987) to verify semantic separation among discourse groups. Clustering performance was evaluated using silhouette coefficients, yielding an average score of 0.11, suggesting limited separation between clusters. This reflects the inherent overlap in real-world discourse data, where thematic boundaries are not sharply defined. Such results are consistent with prior text mining studies, where semantically related topics often produce overlapping cluster structures.

(3) Sentiment classification was cross validated against manual coding of a random sample ($n = 200$) to verify alignment between automated and human interpretations.

(4) Intercoder reliability was assessed using Cohen's kappa, yielding a value of 0.73, indicating substantial agreement between automated sentiment classification and manual coding. This supports the validity of the sentiment analysis despite the use of a general-purpose lexicon (Cohen, 1960).

Beyond algorithmic metrics, a human-in-the-loop validation step was included: three domain experts independently reviewed sample outputs from topic modelling and sentiment clustering, reconciling discrepancies through consensus. This hybrid validation aligns with emerging DSS practice that privileges interpretability and contextual accuracy over black-box precision (Marques and Ferreira, 2020; Król and Zdonek, 2022).

Together, these validation procedures meet DSS expectations for analytic credibility, transparency, and traceability (Burstein and Holsapple, 2008). They ensure that the computational outputs reflect authentic lexical and affective framing of ESG issues in infrastructure projects rather than artifacts of algorithmic bias. This hybrid validation approach also compensates for the limitations of both LDA and lexicon-based sentiment analysis, ensuring robustness while maintaining interpretability.

4. Data analysis

4.1 Analytical approach

The dataset analyzed in this study comprises a structured corpus derived from 17 semi-structured interview transcripts conducted with infrastructure professionals in Australia. Each interview, lasting between 45 min and one hour, was conducted via Zoom, recorded, and transcribed using automated transcription tools, ensuring timely and accurate data capture (77769 words). The interviews followed a standardized protocol designed to elicit insights into the role of digital technologies in enhancing ESG performance, with a particular focus on themes such as automation and structuring of ESG data, data quality and reliability, transparency in reporting, and the enablers and barriers to digital adoption.

From these interview scripts, a total of 122 clusters were generated through computational processing, representing semantically related segments of discourse across operational, managerial, and governance contexts within infrastructure projects. Using advanced NLP techniques, including topic modelling and semantic clustering, 49 distinct topics were subsequently extracted, capturing the dominant thematic structures embedded within the dataset. The corpus reflects a diverse range of communication forms, including narrative responses, reflective insights, and technically oriented discussions, providing a rich linguistic basis for analyzing ESG-related discourse. Purposive sampling was employed to ensure that all participants possessed relevant expertise in infrastructure and digital transformation, thereby strengthening the analytical relevance and interpretive validity of the dataset.

The data comprised practitioner communications from real infrastructure projects, representing multidisciplinary exchanges across technical, managerial, and governance domains. After standard pre-processing (tokenisation, lemmatisation, removal of stop words, and project boilerplate), we applied a two-stage hybrid analysis.

First, topic modelling generated latent themes using unsupervised inference.

Second, semantic clustering on document embeddings consolidated topics into interpretable clusters. This hybrid pipeline mitigated the known instability and redundancy of LDA outputs and improved interpretability; this best practice is increasingly recognised in DSS research for combining machine-driven discovery with human validation (Maibaum *et al.*, 2024; Król and Zdonek, 2022).

Each cluster was coded along five dimensions:

(1) Dominant topic,

(2) ESG dimension (Environmental, Social, or Governance),

(3) Sentiment polarity,

(4) Presence of digital framing (e.g. references to dashboards, sensors, AI, automation), and

(5) framing summary.

Sentiment was derived from a lexicon-based approach and manually cross-checked for domain-specific terminology (e.g. risk, audit, mitigation). Digital framing was identified through keyword heuristics and verified through coder triangulation. The threshold for digital mediation was set at 20% keyword presence within cluster documents, ensuring consistent and replicable classification. This human-in-the-loop refinement ensured interpretability, aligning with current DSS literature on explainable and human-augmented text analytics (Shollo and Galliers, 2016; Arnott and Pervan, 2016; Power, 2022).

In addition to qualitative interpretation, comparative descriptive statistics were computed across ESG dimensions, including mean sentiment scores and standard deviations based on VADER compound sentiment values. To assess whether observed differences were statistically significant, the Kruskal–Wallis test was conducted, given the non-normal distribution of sentiment data across groups. The results indicate a statistically significant difference in sentiment across Environmental, Social, and Governance dimensions ($H = 11.482$, $p = 0.00321$), confirming the presence of sentiment asymmetry in ESG discourse.

Each cluster was manually mapped to Environmental, Social, or Governance dimensions based on dominant lexical themes and keyword distributions, supported by the keyword-per-topic analysis. Topic importance within each cluster was operationalized using frequency of occurrence, ensuring that dominant topics reflect recurrent discourse patterns rather than subjective weighting.

4.2 Top 10 interpretable clusters

From 122 initial clusters, the ten most interpretable and decision-relevant clusters were selected using a multi-criteria filtering approach (See Table 6). Selection criteria included:

(1) *Topic coherence*, ensuring semantic consistency within clusters (Röder *et al.*, 2015);

(2) *Cluster size*, prioritizing clusters with sufficient document representation;

(3) *Silhouette score*, indicating clear separation from other clusters (Rousseeuw, 1987);

(4) *Interpretability*, assessed through expert review based on clarity, relevance to ESG dimensions, and decision-support usefulness.

Table 6. The top 10 interpretable clusters for the findings section

Cluster	Dominant topic	Top keywords	ESG dimension	Sentiment	Digital framing?	Framing summary
2	Carbon Metrics	carbon, emissions, reduction, mitigation, targets, sustainability	Environmental	Neutral	No	"Operational carbon accountability with cautious implementation framing"
4	Social Risk	community, displacement, transport, stakeholder, impact, equity	Social	Neutral	No	"Social vulnerability and stakeholder sensitivity framing"
6	Audit and Reporting	audit, reporting, governance, compliance, monitoring, accountability	Governance	Neutral	No	"Procedural accountability and compliance-oriented governance framing"
12	Data-Driven Compliance	reporting, compliance, dashboard, analytics, digital tools, real-time data	Governance	Positive	Yes	"Digitally enabled compliance and data-centric governance framing"
17	Energy Efficiency	energy, efficiency, optimization, sensors, monitoring, smart systems	Environmental	Positive	Yes	"Technology-driven sustainability optimization framing"
21	Labour Ethics	labour, workforce, standards, safety, violations, worker rights	Social	Negative	No	"Ethical risk and workforce vulnerability framing"
30	Emissions Monitoring	emissions, carbon tracking, monitoring, dashboard, live analytics, reporting	Environmental	Positive	Yes	"Real-time environmental intelligence and performance framing"
35	Governance Automation	automation, audit, compliance, digital governance, workflow, oversight	Governance	Positive	Yes	"Automated governance and algorithmic accountability framing"
42	Community Engagement	community, engagement, trust, participation, dialogue, inclusion	Social	Positive	No	"Relational legitimacy and participatory governance framing"
56	AI and ESG Integration	AI, predictive analytics, risk modelling, scenario analysis, decision support	Governance	Positive	Yes	"AI-enabled strategic foresight and predictive governance framing"

This structured selection process ensures that the reported clusters are not only statistically robust but also substantively meaningful for governance analysis (6). These clusters illustrate the lexical breadth of ESG discourse and highlight how different dimensions are framed linguistically and sentimentally.

This interpretive condensation step aligns with DSS principles of knowledge reduction and managerial insight extraction from unstructured data, converting large-scale text into actionable evidence (Burstein and Holsapple, 2008; Alaei *et al.*, 2019; van der Aalst *et al.*, 2023).

4.3 Patterns by ESG dimension

At an aggregate level, Environmental and Governance clusters exhibited higher mean sentiment scores (E = 0.198, G = 0.234) compared to Social clusters (S = 0.183). Statistical testing using the Kruskal–Wallis test confirmed that these differences are statistically significant at conventional levels (H = 11.48, p = 0.003), supporting the presence of asymmetry in ESG framing.

4.3.1 Environmental. Environmental clusters showed contrasting farming. Carbon Metrics (Cluster 2) displayed neutral sentiment and lacked digital mediation, reflecting a neutral evaluative tone with limited digital framing. In contrast, Energy Efficiency (Cluster 17) and Emissions Monitoring (Cluster 30) were positively framed and explicitly linked to digital dashboards and tracking systems. This pattern suggests that digital instrumentation can foster confidence and innovation-oriented rhetoric, echoing DSS research on how performance dashboards reconfigure managerial perception and sustainability sensemaking (Maibaum *et al.*, 2024; Janssen *et al.*, 2017; Król and Zdonek, 2022).

4.3.2 Social. Social clusters showed a diverse mix of neutral, negative, and positive framings, with limited digital mediation. Social Risk (Cluster 4) used cautious language around supply-chain uncertainties and compliance. Labour Ethics (Cluster 21) exhibited negative sentiment dominated by violation, audit, and sanction. Community Engagement (Cluster 42) adopted a positive tone focused on trust, dialogue, and stakeholder collaboration, but lacked digital references.

This suggests that social issues are framed relationally and normatively rather than technologically, resulting in weaker integration into data-driven reporting (Lee *et al.*, 2025; van der Aalst *et al.*, 2023). From a DSS perspective, this represents a potential representational imbalance, where social values are less codified and harder to translate into machine-readable governance indicators.

4.3.3 Governance. Governance topics demonstrated the strongest digital mediation. Audit and Reporting (Cluster 6) maintained a procedural tone, whereas Data-Driven Compliance (Cluster 12), Governance Automation (Cluster 35), and AI and ESG Integration (Cluster 56) were positively framed and densely populated with digital keywords. Terms such as automation, system, dashboard, and integration indicate the adoption of predictive and automated compliance systems. This aligns with DSS evidence that governance is often the earliest domain to adopt decision automation, AI auditing, and rule-based analytics (Power, 2022; Maibaum *et al.*, 2024; Marques and Ferreira, 2020).

4.3.4 Digital transformation as a framing device. Across the dataset, clusters containing digital keywords were consistently associated with positive sentiment, leading to what we term a “data-driven innovation” framing. In contrast, non-digitally mediated clusters reflected “resistance and scepticism,” particularly within social and environmental discussions. This polarity supports the argument that digital infrastructures act not merely as tools but as framing mechanisms that shape how ESG success is narrated. This is consistent with former research stating that data artefacts structure organizational cognition (Marques and Ferreira, 2020; Shollo and Galliers, 2016; van der Aalst *et al.*, 2023).

4.3.5 Cross-cluster/cross-dimension insights. From the aggregate analysis, three cross-dimensional insights emerged:

(1) Digital framing correlates with positive tone, especially in Environmental and Governance dimensions.

(2) Social themes are less digitally mediated; characterized by cautious or normative language rather than data-driven optimism.

(3) Governance emerges as the most automation-ready domain, confirming DSS studies that identify compliance and auditing as primary entry points for intelligent decision-support tools (Power, 2022; Maibaum *et al.*, 2024).

These patterns align with recent DSS findings that digital artefacts (e.g. dashboards, compliance platforms) shift sustainability discourse toward optimistic, innovation-oriented framing, while non-digitized domains risk underrepresentation in data-driven reporting systems (Lee *et al.*, 2025; Alaei *et al.*, 2019). Likewise, ESG-KIBERT analysis demonstrates that fine-grained linguistic framing (beyond keyword frequency) significantly enhances the decision relevance of sustainability evaluations (Lee *et al.*, 2025). Together, these results indicate that data visibility and rhetorical positivity are positively associated across ESG discourse dimensions, consistent with recent literature on how digital artefacts shape the framing of ESG legitimacy (Lee *et al.*, 2025; Król and Zdonek, 2022).

5. Findings

This section presents the analytical outcomes of the ESG discourse dataset. The results integrate topic clustering, sentiment classification, and lexical framing to address the four research questions. The analysis identifies thematic patterns and examines how digital mediation and linguistic framing are reflected in ESG communication at the discourse level.

5.1 Descriptive findings: answers to research questions

5.1.1 RQ1: What topics and themes emerge most frequently across ESG discourse?. From 122 clusters generated by the hybrid NLP–simheuristic pipeline, the ten most interpretable clusters were identified. Each cluster was labelled based on its dominant theme (e.g. Carbon Metrics, Labour Ethics, and Governance Automation). These clusters reflected recurrent ESG concerns, cutting across environmental, social, and governance domains. As shown in Table 6, these clusters provide the basis for subsequent ESG classification and comparative analysis across dimensions.

The most frequent topics involved digital compliance tools, emissions tracking, and labour standards. Environmental and governance-related clusters were strongly associated with data systems and quantification, while social themes were more relational, human-centred, and normative in tone. This distribution echoes prior observations that environmental metrics are more systematically institutionalized than social ones (Aigbavboa *et al.*, 2024; Maibaum *et al.*, 2024; van der Aalst *et al.*, 2023).

5.1.2 RQ2: How are ESG priorities emphasized or downplayed through language?. Sentiment analysis across clusters revealed differentiated framing patterns. Clusters linked to environmental and governance topics displayed positive sentiment, particularly when associated with digital or data-driven solutions such as dashboards, AI, or automation. In contrast, Social-related discourse exhibits comparatively lower mean sentiment, indicating a less positive evaluative tone relative to Environmental and Governance dimensions. Cluster-level framing summaries indicate variation in evaluative tone and thematic emphasis across ESG categories; this pattern is consistent with the statistically significant differences in sentiment across ESG categories ($H = 11.482$, $p = 0.00321$) reported using the Kruskal–Wallis test. Clusters associated with Environmental and Governance topics exhibit higher mean sentiment scores ($E = 0.198$, $G = 0.234$) compared to social clusters ($S = 0.183$), indicating a relatively more positive evaluative tone in these dimensions. Conversely, Social clusters show lower mean sentiment and more heterogeneous sentiment distribution, consistent with more cautious or mixed framing patterns.

These findings suggest that sentiment toward ESG is not uniform but context-dependent, aligning with decision-support literature that emphasizes interpretive and context-aware analytics (Król and Zdonek, 2022; Roufosse *et al.*, 2024).

5.1.3 RQ3: Are there discernible differences in lexical framing between E, S, and G components?. Lexical frequency and co-occurrence analysis revealed clear distinctions across ESG pillars:

(1) Environmental clusters emphasized efficiency, tracking, and carbon, indicating technical and quantifiable framing.

(2) Social clusters used terms such as community, labour, and trust, suggesting relational and ethical connotations.

(3) Governance clusters foregrounded audit, compliance, and system, pointing to procedural and administrative framing.

These distinctions are not merely linguistic but conceptual, suggesting that each ESG pillar embodies a distinct communication logic (technical, ethical, and procedural, respectively). Such lexical asymmetry has implications for the design of ESG analytics and dashboards that rely on automated text interpretation (Kiriu and Nozaki, 2020; Lee *et al.*, 2025).

5.1.4 RQ4: How do digital technologies get positioned in relation to ESG themes?. The "Digital Framing" dimension, derived from cluster metadata, revealed that digital transformation was primarily positioned as an enabler of ESG performance (especially in the governance and environmental clusters). References to dashboards, AI-based monitoring, and automation were often coupled with positive sentiment. Conversely, social clusters lacked digital mediation, adopting a more analogue and humanistic tone.

This pattern implies a selective coupling between digitalization and ESG performance narratives. It reinforces the notion that digital technologies are perceived as amplifiers of accountability and transparency but not necessarily of social responsibility. These insights parallel recent DSS perspectives on digital sustainability analytics (Maibaum *et al.*, 2024; Power, 2022) and highlight how system design may reproduce rhetorical biases through selective visibility.

5.2 ESG comparative summary

Table 7 summarizes comparative linguistic patterns across ESG dimensions, integrating frequency, sentiment, and digital mediation indicators.

Overall, Governance content is the most prevalent and exhibits the highest average sentiment, while Environmental and Social dimensions appear less frequently and are associated with comparatively lower sentiment scores.

This comparative structure enables systematic interpretation of ESG asymmetries beyond topic frequency, highlighting differences in sentiment, digital mediation, and framing logic across dimensions. Tables 1–4 is not presented as descriptive outputs only, but as integrated analytical components that support cross-dimensional ESG comparison across thematic, sentiment, and digital framing structures.

5.3 Comparative insights: positioning against prior research

To complement the interpretive findings, descriptive statistics were computed across ESG dimensions. Environmental and Governance clusters exhibit higher mean sentiment scores ($E = 0.198$, $G = 0.234$) than social clusters ($S = 0.183$). A Kruskal–Wallis test indicates that these differences are statistically significant ($H = 11.482$, $p = 0.00321$), demonstrating systematic variation in evaluative tone across discourse categories. The corresponding effect size is small ($\varepsilon^2 \approx 0.009$), suggesting that, although reliable, the magnitude of these differences is limited. Our analysis identifies three overarching patterns:

Table 7. ESG dimensions

ESG dimension	Avg sentiment	Std Dev	Frequency (%)	Dominant framing
Environmental	0.198	0.198	0.198	Technical/Data-driven
Social	0.183	0.353	16.55	Relational/Normative
Governance	0.234	0.355	63.52	Procedural/Control

5.3.1 Variation across ESG dimensions. Prior ESG text-mining research often indicates that environmental issues dominate corporate disclosure due to their quantifiability and alignment with measurable KPIs (Aigbavboa *et al.*, 2024; Maibaum *et al.*, 2024). In contrast, our results indicate that Governance-related discourse is the most prevalent, while Environmental and Social topics appear less frequently. At the same time, Environmental and Governance clusters exhibit higher average sentiment than social clusters, and these differences are statistically significant (H = 11.482, p = 0.00321), albeit with a small effect size. This points to a measurable but modest asymmetry in evaluative tone across ESG dimensions.

5.3.2 Framing and sentiment patterns. Clusters referencing digital tools (dashboards, AI systems, compliance monitoring) exhibited systematically positive sentiment, reinforcing the view that digitalization enhances perceptions of transparency and accountability (Lee *et al.*, 2025). Yet, this also introduces a potential digitization bias, where optimism arises not from substantive ESG outcomes but from the mere presence of technological mediation.

Across clusters, variation in sentiment suggests differences in how ESG topics are discursively framed. Environmental and Governance themes tend to be associated with relatively more positive evaluative language, whereas social topics display comparatively lower sentiment. These differences, while statistically reliable, remain limited in magnitude, indicating that sentiment variation should be interpreted cautiously and in conjunction with qualitative framing analysis.

5.3.3 Relative under-representation of social themes. The lower frequency and comparatively weaker sentiment of Social topics suggest that this dimension occupies a less prominent position within the overall discourse. This finding aligns with prior research emphasizing the challenges of representing socially oriented issues, which are often more qualitative and context-dependent (Unerman and Chapman, 2014). It also highlights a potential limitation of approaches that rely heavily on frequency or sentiment metrics, as these may understate the complexity and nuance of Social-related content.

5.3.4 Integrative summary of analytical insights. Overall, while prior NLP applications in ESG research have focused on automated classification and scoring (Kiriu and Nozaki, 2020; Roufosse *et al.*, 2024), this study highlights the value of examining sentiment variation across ESG dimensions. The findings demonstrate that statistically significant—though small—differences in evaluative tone exist, with Governance and Environmental discourse tending to be more positive than social discourse. This supports a more nuanced, interpretive perspective on ESG communication, where quantitative indicators are complemented by attention to how language shapes the representation of sustainability issues.

6. Discussion

6.1 Theoretical contributions

This study advances ESG discourse, digital framing, and decision support in three main ways.

First, it positions lexical framing as a measurable governance dimension. Many computational ESG studies extract topic frequencies or ESG scores (Kiriu and Nozaki, 2020; Roufosse *et al.*, 2024), but we treat how topics are framed—through sentiment, digital cues, and narrative tone—as governance-relevant metadata. Each cluster receives a framing summary (e.g. data-driven innovation, resistance and scepticism), serving as interpretable decision signals, extending DSS research toward semantic explainability (Lee *et al.*, 2025; Maibaum *et al.*, 2024; Król and Zdonek, 2022).

Second, we address the invisibility of Social ESG. Social discourse shows minimal digital mediation, so dashboards rely solely on metrics, which may under-represent relational and qualitative concerns inherent in social sustainability discourse. Managers should integrate unstructured qualitative data (emails, meetings, stakeholder dialogues) into DSS pipelines to balance numeric and discursive inputs (Unerman and Chapman, 2014; Maibaum *et al.*, 2024).

Third, human-in-the-loop interpretability is maintained. Automated framing functions as advisory cues: users inspect text, correct misclassifications, and iteratively retrain models, supporting adaptive, transparent DSS aligned with Responsible Digital Transformation principles (van der Aalst *et al.*, 2023; Power, 2022).

Fourth, self-reinforcing feedback loops are prevented. Positive discourse may be influenced by the presence of digital monitoring systems rather than solely reflecting substantive changes in underlying ESG performance indicators; periodic audits of non-instrumented domains mitigate systemic bias.

It is important to emphasize that the analysis remains strictly at the linguistic and discourse level. No inference is made regarding managerial cognition, behavioural decision-making, or actual governance outcomes. Instead, the study identifies structured patterns of ESG framing that may inform decision-support interpretation. Finally, framing diagnostics informs governance action. Persistent absence of “innovation” framing in social topics should trigger strategic review or resource reallocation, operationalizing interpretive analytics as a core DSS governance capability (Lee *et al.*, 2025; Maibaum *et al.*, 2024).

6.2 Practical and managerial implications

This study offers various practical implications for infrastructure project governance and decision-support systems. Most importantly, it demonstrates how ESG-related communication can be transformed into actionable insights through a framing-informed dashboard to support more transparent and reflective decision-making.

Framing dashboards should output from NLP analysis into a set of interpretable indicators. Rather than presenting only traditional ESG metrics, the dashboard would include (1) sentiment signals indicating whether sustainability issues are framed positively, neutrally, or negatively; (2) topic salience showing which ESG dimensions (environmental, social, governance) receive attention; (3) framing asymmetries, highlighting over- or under-emphasized themes; and (4) digital vocabulary signals, identifying when technological language (e.g. “AI”, “dashboard”, “tracking”) is used as a proxy for sustainability claims. These elements could be visualized through trend lines, heat maps, or alert indicators embedded within existing project reporting systems.

In practice, project managers could use such a dashboard to complement traditional performance data. For example, if environmental topics are consistently framed positively while social concerns appear only in risk-oriented language, this may signal an imbalance in how sustainability priorities are interpreted and communicated. Similarly, frequent references to digital tools without corresponding discussion of outcomes may indicate symbolic compliance or “technological optimism.” These signals would not replace managerial judgement but act as prompts for reflection, discussion, and corrective action during project reviews.

From a governance perspective, this approach can support greater accountability by making communication patterns visible. By identifying discrepancies between reported achievements and underlying discourse, organizations may be better positioned to detect potential greenwashing or overly optimistic reporting. For instance, if sustainability claims are consistently associated with promotional language but lack evidence of trade-offs or constraints, governance bodies can initiate deeper scrutiny. In this way, framing diagnostics provide an additional layer of oversight beyond conventional ESG metrics.

However, these benefits must be interpreted with caution. NLP-based insights are inherently probabilistic and sensitive to model design, data quality, and contextual interpretation. There is a risk of misinterpretation if framing signals are treated as objective indicators rather than heuristic cues. For example, negative sentiment may reflect legitimate risk management rather than poor performance, while technical language may be necessary in certain project phases. Therefore, framing analytics should be used as a decision-support tool rather than a decision-making substitute, supported by human validation and contextual expertise.

Importantly, the practical implications of this study remain exploratory and conceptually oriented. While the proposed dashboard illustrates how framing diagnostics could be operationalized, further empirical validation is required to assess its effectiveness across different project settings. Future research could test how such tools influence managerial decisions, stakeholder engagement, and governance outcomes in real-world infrastructure projects.

6.3 Limitations

This study has several limitations that should be acknowledged.

First, the study adopts a cross-sectional design to capture structural patterns of ESG lexical framing at a representative point in time. The objective is not to model temporal dynamics, but to identify systematic differences in how ESG dimensions and digital transformation are linguistically framed within project communications. While this enables controlled comparison across ESG categories, it limits the ability to capture how ESG discourse evolves.

Second, lexicon-based sentiment analysis (VADER) provides transparent and reproducible scoring but is not tailored to domain-specific ESG language (Hutto and Gilbert, 2014). Technical and governance-related terms such as "risk," "audit," or "compliance" may carry neutral or procedural meanings in infrastructure contexts, yet may be assigned negative polarity by general-purpose lexicons. Although manual validation and expert review were applied to mitigate this limitation, some degree of domain misclassification may persist. In addition, intercoder reliability statistics were not formally reported, which may limit the robustness of validation.

Third, unsupervised learning methods such as LDA and k-means are sensitive to term frequency distributions and may reflect dominant linguistic patterns in the dataset (Blei *et al.*, 2003; Pedregosa *et al.*, 2011). As a result, emerging or less frequent ESG themes may be under-represented, introducing potential algorithmic bias in topic and cluster formation. Fourth, the dataset reflects formal project communications and therefore privileges institutional and managerial voices. Informal stakeholder perspectives, community narratives, and marginalized actors may be underrepresented or absent. This introduces representational bias, as the analysis captures documented organizational discourse rather than the full spectrum of ESG-relevant perspectives. Fifth, digital framing detection relies on a predefined keyword dictionary and contextual filtering. Although ambiguity was addressed through manual validation and false-positive checks, keyword-based approaches may still introduce classification bias due to polysemy and contextual variation in language use.

Finally, the findings are primarily sector-specific to infrastructure project environments, where governance structures and technical complexity shape ESG discourse. While conceptual insights on digital mediation and lexical framing may be transferable to other data-intensive sectors, generalization should be made cautiously and validated empirically. This enables DSS to move beyond outcome-oriented ESG scoring toward diagnostic capabilities that identify discursive asymmetries, visibility gaps, and framing biases across ESG dimensions—capabilities not captured by existing ESG text classification or sentiment analysis approaches.

6.4 Future research directions

This study opens several avenues for future research.

First, longitudinal or intervention-based studies could examine how ESG framing evolves, particularly in response to policy changes, digital adoption, or external shocks. This would provide deeper insight into the dynamics of ESG discourse beyond cross-sectional analysis. Second, replication across sectors such as healthcare, energy, and supply chains would help assess whether the observed relationship between digital mediation and ESG framing is context-specific or more broadly generalizable. Third, future research could extend the computational approach by incorporating advanced models capable of detecting mixed sentiment, hedging, and implicit evaluative tone, complementing lexicon-based pipelines and enhancing interpretive depth. Fourth, experimental or field-based studies could evaluate how framing-aware dashboards influence decision-support processes, including risk identification, governance deliberation, and sustainability assessment. Finally, extending the analysis to cross-cultural and multilingual ESG discourse represents a promising direction. Infrastructure projects operate across diverse institutional and linguistic contexts, where ESG framing may vary significantly. Multilingual NLP models and comparative discourse analysis could provide deeper insight into how cultural and regulatory environments shape ESG communication and digital framing patterns.

7. Conclusion

This study clarifies how ESG discourse is constructed, mediated, and operationalized in infrastructure projects. Environmental and Governance themes are tightly linked to digital infrastructures and framed optimistically, while social dimensions remain relational, ethical, and cautious. We provide a scalable, explainable computational framework transforming domain discourse into decision-relevant framing indicators, extending text analytics from disclosure-level reporting to embedded project communication. By establishing a lexical framing taxonomy rooted in practitioner practice, we offer a diagnostic mechanism to identify accentuated, marginalized, or rhetorically constrained ESG elements. Embedding framing analytics into DSS architecture enhances transparency, interpretability, and governance reflexivity in sustainable infrastructure delivery.

Concretely, the findings imply that infrastructure project managers should integrate framing-based alerts into routine project reviews, using discrepancies between positive environmental discourse and risk-oriented social narratives as triggers for rebalancing attention and resource allocation. For ESG policy designers, the identified digitization bias highlights the need to formalize requirements for incorporating qualitative and unstructured social data, such as stakeholder engagement records, into ESG reporting frameworks, ensuring that relational dimensions are not systematically excluded from data-driven governance. For developers of decision support systems, the proposed lexical framing taxonomy can be operationalized as a “decision metadata” layer, where each ESG indicator is augmented with sentiment, framing category, and digital signal tags, enabling more interpretable and reflexive analytics. Embedding such metadata alongside conventional metrics allows systems not only to report performance, but also to reveal how sustainability is being constructed and communicated in practice. In this way, addressing digitization bias becomes a design and governance priority, supporting more socially balanced and context-sensitive sustainable infrastructure delivery.

Ethics approval

Human ethical approval was obtained from the second author’s affiliated institute.

References

Abdel-Tawab, S., Abdel-Rahman, M. and El-Bastawisy, A.M. (2023), “The role of BIM in promoting environmental sustainability in construction projects”, *Buildings*, Vol. 13 No. 3, p. 758, doi: 10.3390/buildings13030758.

Aigbavboa, C., Ebekozien, A., Afetorgbor, E.K., Arthur-Aidoo, B.M. and Thwala, W.D. (2024), “Investigating the barriers facing corporate social responsibility in the built environment: Ghana’s perspective from a qualitative approach”, *Property Management*, Vol. 42 No. 4, pp. 493-506, doi: 10.1108/PM-07-2023-0060.

Akomea-Frimpong, I., Ofori, J.N.A., Botchway, B., Tumpa, R.J. and Tetteh, P.A. (2025), “Key success factors towards circular economy in public-private partnership projects”, *Smart and Sustainable Built Environment*, pp. 1-27, doi: 10.1108/SASBE-02-2025-0074.

Alaei, A.R., Becken, S. and Stantic, B. (2019), "Sentiment analysis in tourism: capitalizing on big data", *Journal of Travel Research*, Vol. 58 No. 2, pp. 175-191, doi: 10.1177/0047287517747753.

Ansar, A., Flyvbjerg, B., Budzier, A. and Lunn, D. (2016), "Big projects and big decisions: overestimating benefits and underestimating risks in mega-projects", *Oxford Review of Economic Policy*, Vol. 32 No. 3, pp. 411-427, doi: 10.1093/oxrep/grw020.

Arnott, D. and Pervan, G. (2016), "A critical analysis of decision support systems research revisited: the rise of design science", *Decision Support Systems*, Vol. 86, pp. 84-95, doi: 10.1016/j.dss.2016.03.003.

Baldini, M., Maso, L.D., Liberatore, G., Mazzi, F. and Terzani, S. (2018), "Role of country-and firm-level determinants in environmental, social, and governance disclosure", *Journal of Business Ethics*, Vol. 150 No. 1, pp. 79-98, doi: 10.1007/s10551-016-3139-1.

Benford, R.D. and Snow, D.A. (2000), "Framing processes and social movements: an overview and assessment", *Annual Review of Sociology*, Vol. 26 No. 1, pp. 611-639, doi: 10.1146/annurev.soc.26.1.611.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Sebastopol, CA.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

Boiral, O. (2009), "Greening the corporation through organizational citizenship behaviors", *Journal of Business Ethics*, Vol. 87 No. 2, pp. 221-236, doi: 10.1007/s10551-008-9881-2.

Boiral, O. and Henri, J.-F. (2017), "Is sustainability performance comparable? A study of consistency in GRI reporting", *Business & Society*, Vol. 56 No. 2, pp. 283-317.

Brachem, J. and Rothe, A. (2021), "Stop removing stop words: an evaluation of preprocessing techniques for twitter sentiment analysis with a deep learning approach", *Learning Deep Textwork*, Vol. 37.

Burstein, F. and Holsapple, C.W. (2008), *Handbook on Decision Support Systems 2: Variations*, Springer Science and Business Media, doi: 10.1007/978-3-540-48713-5.

Chan, A.P.C., Yeung, J.F.Y. and Zhao, X. (2022), "Managing geopolitical risks in belt and Road infrastructure projects", *International Journal of Project Management*, Vol. 40 No. 6, pp. 576-589, doi: 10.1016/j.ijproman.2022.04.004.

Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37-46, doi: 10.1177/001316446002000104.

Compact, U.G. (2006), *UN Global Compact*.

Cornelissen, J.P., Durand, R., Fiss, P.C., Lammers, J.C. and Vaara, E. (2015), "Putting communication front and center in institutional theory and analysis", *Academy of Management Review*, Vol. 40 No. 1, pp. 10-27, doi: 10.5465/amr.2014.0381.

Debrah, C., Chan, A.P.C., Darko, A. and Owusu-Manu, D. (2022), "Barriers to sustainable construction in developing countries: a review", *Journal of Cleaner Production*, Vol. 332, 129823, doi: 10.1016/j.jclepro.2021.129823.

Del Gesso, C. and Lodhi, R.N. (2025), "Theories underlying environmental, social and governance (ESG) disclosure: a systematic review of accounting studies", *Journal of Accounting Literature*, Vol. 47 No. 2, pp. 433-461, doi: 10.1108/JAL-08-2023-0143.

Eccles, R.G., Lee, L.E. and Stroehle, J.C. (2020), "The social origins of ESG: an analysis of Innovest and KLD", *Organization & Environment*, Vol. 33 No. 4, pp. 575-596.

Entman, R.M. (1993), "Framing: toward clarification of a fractured paradigm", *Journal of Communication*, Vol. 43 No. 4, pp. 51-58, doi: 10.1111/j.1460-2466.1993.tb01304.x.

Flyvbjerg, B. (2014), "What you should know about megaprojects and why: an overview", *Project Management Journal*, Vol. 45 No. 2, pp. 6-19, doi: 10.1002/pmj.21409.

Gałecka-Drozda, A., Wilkaniec, A., Szczepańska, M. and Świerk, D. (2021), "Potential nature-based solutions and greenwashing to generate green spaces: developers' claims versus reality in new housing offers", *Urban Forestry and Urban Greening*, Vol. 65, 127345, doi: 10.1016/j.ufug.2021.127345.

Gao, Q., Wang, J. and Zhang, R. (2025), "Customer-driven supply chain sustainability in the digital economy: identifying the role of customer firms' digital transformation on supplier ESG performance in China", *International Journal of Productivity and Performance Management*, Vol. 74 No. 8, pp. 2837-2856, doi: 10.1108/IJPPM-11-2024-0799.

Gasperini, A. (2019), "Principles for responsible investment (PRI) and ESG factors", in *Climate Action*, Springer International Publishing, Cham, pp. 737-749.

Goh, C.S., Jack, L. and Bajracharya, A. (2020), "Qualitative study of sustainability policies and guidelines in the built environment", *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, Vol. 12 No. 2, 04520016, doi: 10.1061/(ASCE)LA.1943-4170.0000395.

Gong, J., Yang, R.J. and Wang, Y. (2024), "ESG integration challenges in infrastructure investment: a systematic review", *Sustainable Cities and Society*, Vol. 104, 105286, doi: 10.1016/j.scs.2024.105286.

Honnibal, M. and Montani, I. (2017), "spaCy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing", *SoftwareX*, Vol. 6, pp. 110-118, doi: 10.21203/rs.3.rs-3350837/v1.

Hosseini, M.R., Najafi, M. and Rahimian, F. (2026), "Digital transformation and smart technologies for sustainable built environment: innovations, applications and future pathways", *Smart and Sustainable Built Environment*, Vol. 15 No. 1, pp. 1-8, doi: 10.1108/SASBE-02-2026-0689.

Hutto, C. and Gilbert, E. (2014), "VADER: a parsimonious rule-based model for sentiment analysis of social media text", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8 No. 1, pp. 216-225, doi: 10.1609/icwsm.v8i1.14550.

Jain, A.K. (2010), "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, Vol. 31 No. 8, pp. 651-666, doi: 10.1016/j.patrec.2009.09.011.

Janssen, M., Van Der Voort, H. and Wahyudi, A. (2017), "Factors influencing big data decision-making quality", *Journal of Business Research*, Vol. 70, pp. 338-345, doi: 10.1016/j.jbusres.2016.08.007.

Kallioupi, R.N. and Triantafyllou, S. (2025), "ESG marketing meets sports journalism: media framing and public perception of corporate responsibility in athletic sponsorships", *British Journal of Marketing Studies*, Vol. 13 No. 4, pp. 31-52, doi: 10.37745/bjms.2013/vol13n43152.

Kiriu, T. and Nozaki, M. (2020), "A text mining model to evaluate firms' ESG activities: an application for Japanese firms", *Asia-Pacific Financial Markets*, Vol. 27 No. 4, pp. 621-632, doi: 10.1007/s10690-020-09309-1.

Kouhizadeh, M. and Sarkis, J. (2018), "Blockchain practices, potentials, and perspectives in greening supply chains", *Sustainability*, Vol. 10 No. 10, p. 3652, doi: 10.3390/su10103652.

Król, K. and Zdonek, D. (2022), "Digital assets in the eyes of generation Z: perceptions, outlooks, concerns", *Journal of Risk and Financial Management*, Vol. 16 No. 1, p. 22, doi: 10.3390/jrfm16010022.

Kwon, W., Clarke, I. and Wodak, R. (2009), "Organizational decision-making, discourse, and power: integrating across contexts and scales", *Discourse and Communication*, Vol. 3 No. 3, pp. 273-302, doi: 10.1177/1750481309337208.

Lakhal, N., Aazami, A. and Kummer, S. (2025), "The role of artificial intelligence across the source-to-pay framework: theoretical and practical aspects", *Digital Business*, Vol. 5 No. 2, 100160, doi: 10.1016/j.digbus.2025.100160.

Larbi, J.A., Tang, L., Yevu, S.K., Opoku, A. and Ababio, B.K. (2026), "Ecosystem of drivers for adopting integrated digital delivery for sustainable construction", *Smart and Sustainable Built Environment*, Vol. 15 No. 2, pp. 807-830, doi: 10.1108/SASBE-08-2024-0320.

Lee, H., Kim, J.H. and Jung, H.S. (2025), "ESG-KIBERT: a new paradigm in ESG evaluation using NLP and industry-specific customization", *Decision Support Systems*, Vol. 193, 114440, doi: 10.1016/j.dss.2025.114440.

Li, Y., Zhang, X. and Liu, M. (2021), "ESG capability and performance in construction enterprises", *Journal of Cleaner Production*, Vol. 320, 128917, doi: 10.1016/j.jclepro.2021.128917.

Maibaum, F., Kriebel, J. and Foege, J.N. (2024), "Selecting textual analysis tools to classify sustainability information in corporate reporting", *Decision Support Systems*, Vol. 183, 114269, doi: 10.1016/j.dss.2024.114269.

Mandičák, T., Spišáková, M. and Mésároš, P. (2024), "Sustainable design and building information modeling of construction project management towards a circular economy", *Sustainability*, Vol. 16 No. 11, p. 4376, doi: 10.3390/su16114376.

Marques, P. and Ferreira, F.A.F. (2020), "Digital transformation in decision support systems: a systematic review and future research agenda", *Decision Support Systems*, Vol. 136, 113351, doi: 10.1016/j.dss.2020.113351.

Maurya, K.K. and Biswas, A. (2021), "Performance assessment of governance in Indian smart city development", *Smart and Sustainable Built Environment*, Vol. 10 No. 4, pp. 653-680, doi: 10.1108/SASBE-04-2019-0045.

Müller, R., Pemsel, S. and Shao, J. (2014), "Organizational enablers for governance and governmentality of projects: a literature review", *International Journal of Project Management*, Vol. 32 No. 8, pp. 1309-1320, doi: 10.1016/j.ijproman.2014.07.005.

Müller, R., Zhai, L., Wang, A. and Shao, J. (2016), "A framework for governance of projects: governmentality, governance structure and projectification", *International Journal of Project Management*, Vol. 34 No. 6, pp. 957-969, doi: 10.1016/j.ijproman.2016.05.002.

Najafi, M., Hosseini, M.R. and Rahimian, F. (2025), "Digital frontiers in construction, materials and infrastructures for smart and sustainable built environment", *Smart and Sustainable Built Environment*, Vol. 14 No. 3, pp. 577-581, doi: 10.1108/SASBE-05-2025-0527.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011), "Scikit-learn: machine learning in python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830.

Peng, Y., Chen, H. and Li, T. (2023), "The impact of digital transformation on ESG: a case study of Chinese-listed companies", *Sustainability*, Vol. 15 No. 20, 15072, doi: 10.3390/su152015072.

Perazzoli, S., de Santana Neto, J.P. and de Menezes, M.J.M.B. (2022), "Systematic analysis of constellation-based techniques by using natural language processing", *Technological Forecasting and Social Change*, Vol. 179, 121674, doi: 10.1016/j.techfore.2022.121674.

Poorisat, T., Aigwi, I.E., Doan, D.T. and GhaffarianHoseini, A. (2026), "Unveiling the roles and effective engagement of stakeholders involved in sustainable building design and practices", *Smart and Sustainable Built Environment*, pp. 1-26, doi: 10.1108/SASBE-10-2025-0678.

Power, D.J. (2022), "Explainable AI and decision automation in governance", *Decision Support Systems*, Vol. 158, 113757, doi: 10.1016/j.dss.2022.113757.

Provost, F. and Fawcett, T. (2013), *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O'Reilly Media, Sebastopol, CA.

Qi, Y., Zhang, L. and Wang, J. (2023), "ESG integration in infrastructure projects: challenges and opportunities", *Journal of Cleaner Production*, Vol. 387, 135767, doi: 10.1016/j.jclepro.2023.135767.

Radzi, A.R., Almutairi, S. and Rahman, R.A. (2026), "Environmental, social and governance (ESG) in the built environment: science mapping of present and future trends", *International Journal of Building Pathology and Adaptation*, Vol. 44 No. 5, pp. 1-18, doi: 10.1108/IJBPA-07-2025-0170.

Ramohlokoane, M.M., Awuzie, B.O., Sathvik, S. and Singh, A.K. (2025), “Evaluating sustainability-related deliverables across the construction project lifecycle: a Delphi study”, *Smart and Sustainable Built Environment*. doi: 10.1108/SASBE-10-2023-0299.

Řehůřek, R. and Sojka, P. (2010), “Software framework for topic modelling with large corpora”, *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, University of Malta, pp. 45–50, available at: https://repozitar.cz/publication/15725/?lang=en;kod=S530

Röder, M., Both, A. and Hinneburg, A. (2015), “Exploring the space of topic coherence measures”, *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 399-408, doi: 10.1145/2684822.2685324.

Rose, S., Engel, D., Cramer, N. and Cowley, W. (2010), “Automatic keyword extraction from individual documents”, in Berry, M.W. and Kogan, J. (Eds), *Text Mining: Applications and Theory*, Wiley, pp. 1-20, doi: 10.1002/9780470689646.ch1.

Roufosse, A., Fréour, C. and Loukil, K. (2024), “ESG-KIBERT: knowledge-aware transformers for sustainability classification”, *Decision Support Systems*, Vol. 183, 114269, doi: 10.1016/j.dss.2023.114269.

Rousseeuw, P.J. (1987), “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”, *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65, doi: 10.1016/0377-0427(87)90125-7.

Shang, G., Zeng, S. and Xu, S. (2023), “Knowledge structuring from unstructured text: a hybrid DSS approach”, *Decision Support Systems*, Vol. 167, 113865, doi: 10.1016/j.dss.2023.113865.

Shollo, A. and Galliers, R.D. (2016), “Towards an understanding of the role of business intelligence systems in organizational knowing”, *Information Systems Journal*, Vol. 26 No. 4, pp. 339-367, doi: 10.1111/isj.12071.

Singh, A.K. and Kumar, V.P. (2024), “Establishing the relationship between the strategic factors influencing blockchain technology deployment for achieving SDG and ESG objectives during infrastructure development: an ISM-MICMAC approach”, *Smart and Sustainable Built Environment*, Vol. 13 No. 3, pp. 711-736, doi: 10.1108/SASBE-12-2023-0405.

Su, X., Wang, S. and Li, F. (2023), “The impact of digital transformation on ESG performance based on the mediating effect of dynamic capabilities”, *Sustainability*, Vol. 15 No. 18, 13506, doi: 10.3390/su151813506.

Too, E.G. and Weaver, P. (2014), “The management of project management: a conceptual framework for project governance”, *International Journal of Project Management*, Vol. 32 No. 8, pp. 1382-1394, doi: 10.1016/j.ijproman.2013.07.006.

Tumpa, R.J., Naeni, L.M., Afzal, F. and Ghanbaripour, A.N. (2025), “Leveraging digital technology to improve environmental, social, and governance performance of infrastructure projects”, *Management Decision*, Vol. 63 No. 13, pp. 455-496, doi: 10.1108/MD-04-2024-0818.

Tumpa, R.J., Baker, M. and Ghanbaripour, A.N. (2026), “From diversity to social value: insights from infrastructure project professionals”, *International Journal of Managing Projects in Business*, Vol. 19 No. 2, pp. 302-328, doi: 10.1108/IJMPB-07-2024-0179.

Unerman, J. and Chapman, C. (2014), “Academic contributions to enhancing accounting for sustainable development”, *Accounting, Organizations and Society*, Vol. 39 No. 6, pp. 385-394, doi: 10.1016/j.aos.2014.07.003.

Ure, J.D., South, A.J., Farnsworth, C.B., Bown, M. and Thompson, N. (2024), “Enlarging the role of CEM professionals in corporate sustainability: ESG programs and the built environment”, *Journal of Construction Engineering and Management*, Vol. 150 No. 8, 04024095, doi: 10.1061/JCEMD4.COENG-1360.

van der Aalst, W.M., Hinz, O. and Weinhardt, C. (2023), “Sustainable systems engineering: WMP van der Aalst et al”, *Business and Information Systems Engineering*, Vol. 65 No. 1, pp. 1-6, doi: 10.1007/s12599-022-00784-6.

Van der Velden, M. (2018), “Digitalisation and the UN sustainable development goals: what role for design”, *IxD&A*, Vol. 37 No. 2018, pp. 160-174, doi: 10.55612/s-5002-037-008.

Van Tam, N. and Quoc Toan, N. (2026), "From ESG to value: how green innovation transforms the sustainability trajectory of construction firms", *Engineering, Construction and Architectural Management*, pp. 1-23, doi: 10.1108/ECAM-03-2025-0534.

Villa, C., Olsson, N. and Langford, D. (2023), "ESG integration in project governance: emerging practices and challenges", *International Journal of Project Management*, Vol. 41 No. 2, pp. 189-203, doi: 10.1016/j.ijproman.2022.10.005.

Wang, L., Wang, X. and Liu, X. (2025a), "Project governance and governance of interorganizational project networks: toward understanding their relationships and future research agenda", *Project Management Journal*, Vol. 56 No. 1, pp. 31-40, doi: 10.1177/87569728241234567.

Wang, Y., Lin, W., Yan, Z. and Hu, Y. (2025b), "Mechanism for improving carbon reduction efficiency in the construction industry based on ESG information disclosure", *Engineering, Construction and Architectural Management*, pp. 1-31, ahead-of-print, doi: 10.1108/ECAM-10-2024-1403.

Weerakoon, P. and Perera, B.A.K.S. (2025), "Stakeholder engagement framework to foster circularity throughout the construction project lifecycle: a qualitative Delphi study", *Smart and Sustainable Built Environment*, pp. 1-24, doi: 10.1108/SASBE-05-2025-0227.

Weick, K.E. (1995), *Sensemaking in Organizations*, Sage Publications, Thousand Oaks, CA, Vol. 3, No. 10.1002.

Zwikael, O. and Huemann, M. (2023), "Project benefits management: making an impact on organizations and society through projects and programs", *International Journal of Project Management*, Vol. 41 No. 8, 102538, doi: 10.1016/j.ijproman.2023.102538.

Further reading

Cai, W., Fang, C. and Zhang, Z. (2025), "Evaluating ESG performance in construction enterprises using artificial intelligence", *Journal of Cleaner Production*, Vol. 457, 144974, doi: 10.1016/j.jclepro.2024.144974.

Nervino, E., Wang, J. and Ma, C. (2025), "Self-mentions and stakeholders in climate change discourse: the case of banks", *Languages and Dialogues*, Vol. 15 No. 1, pp. 56-80, doi: 10.1075/ld.00187.ner.

Plec, E. and Pettenger, M. (2012), "Greenwashing, corporate discursive strategies, and their implications for climate change engagement", *Environmental Communication*, Vol. 6 No. 3, pp. 193-211.

Tumpa, R.J. and Naeni, L. (2025), "Improving decision-making and stakeholder engagement at project governance using digital technology for sustainable infrastructure projects", *Smart and Sustainable Built Environment*, Vol. 14 No. 4, pp. 1292-1329, doi: 10.1108/SASBE-10-2024-0451.

Zhang, M. and Huang, Z. (2024), "The impact of digital transformation on ESG performance: the role of supply chain resilience", *Sustainability*, Vol. 16 No. 17, p. 7621, doi: 10.3390/su16177621.

Corresponding author

Mahshid Tootoonchy can be contacted at: m.tootoonchy@gmail.com, mahshid.tootoonchyhamedani@vu.edu.au